Mise en place d’un Data Warehouse: Les titres Netflix et leurs notes sur IMDB

### Nouvembre 2024

**Contents**

* 1. [Introduction](#_bookmark0) 2
  2. [Business Intelligence](#_bookmark1) 2
  3. [Data Warehouse](#_bookmark2) 3
  4. [Id´ee](#_bookmark3) 3
  5. [Conception](#_bookmark4) 4
  6. [R´ealisation](#_bookmark5) 4
     1. [Phase: Collection des donn´ees](#_bookmark6) 5
     2. [Int´egration des donn´ees](#_bookmark7) 6
     3. [Analyse des donn´ees](#_bookmark8) 8
     4. [Reporting](#_bookmark9) 10
  7. [Conclusion](#_bookmark10) 12

# Introduction

Au cours des derni`eres ann´ees, la Business Intelligence passait d’un domaine qui appartenait principalement a` l’IT `a une fonction devenue cl´e dans toutes entreprises, pour d´etecter des opportunit´es business, optimiser ses campagnes marketing ou financi`eres, par exemple. Aujourd’hui, cependant, il existe toute une gamme de solutions de Business Intelligence intuitives et acces- sibles qui permettent aux utilisateurs de toute l’entreprise de se connecter et d’acc´eder aux tableaux de bord BI et aux analyses qui transforment les donn´ees brutes en v´eritables sources d’informations commerciales.

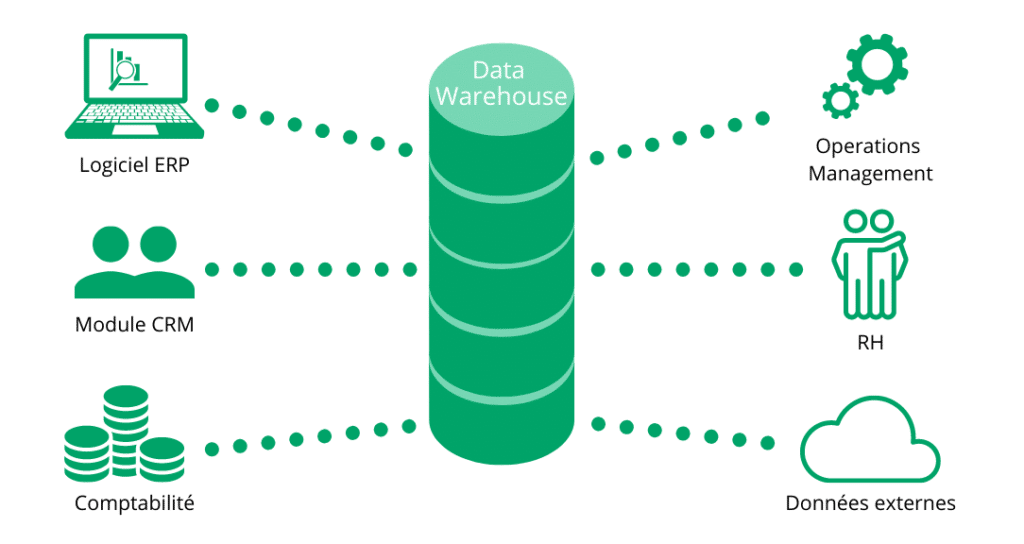
# Business Intelligence

****

La **Business Intelligence (BI)** est un processus technologique d’analyse des donn´ees et de pr´esentation d’informations pour aider les dirigeants, man- agers et autres utilisateurs finaux de l’entreprise `a prendre des d´ecisions busi- ness ´eclair´ees. La Business Intelligence englobe une grande vari´et´e d’outils, d’applications et de m´ethodologies qui permettent aux organisations de col- lecter des donn´ees `a partir de syst`emes internes et de sources externes. Ces donn´ees sont ensuite pr´epar´ees pour l’analyse afin de cr´eer des rapports, tableaux de bord et et autres outils de de Data Viz pour rendre les r´esultats analytiques disponibles aux d´ecideurs et aux op´erations.

Aujourd’hui, les entreprises s’appuient sur les logiciels de Business In- telligence pour identifier et extraire des informations pr´ecieuses des grands volumes de donn´ees qu’elles stockent. Ces outils permettent d’en tirer des informations tels que des veilles concurrentielles et les tendances du march´e, ainsi que des informations internes tel que trouver les raisons des opportu- nit´es perdues.

# Data Warehouse

****

Un entrepˆot de donn´ees (Data Warehouse) est un d´epˆot central con- tenant des informations qui peuvent analys´ees dans le but de prendre des d´ecisions plus ´eclair´ees. Les donn´ees stock´ees dans l’entrepoˆt de donn´ees proviennent de syst`emes transactionnels, de bases de donn´ees relationnelles et d’autres sources, et arrivent g´en´eralement a` intervalle r´egulier. Les ana- lystes d’affaires, les ing´enieurs des donn´ees, les scientifiques des donn´ees et les dirigeants acc`edent aux donn´ees via des outils de Business Intelligence (BI), des clients SQL et d’autres applications d’analyse.

# Id´ee

Notre id´ee est de mettre en place l’entrepoˆt de donn´ees qui s’agit d’analyser Les titres Netflix et leurs notes sur IMDB.

# Conception

Pour faire ¸ca, nous suivrons la m´ethodologie de conception **Kimball**.

Initi´e par Ralph Kimball, le mod`ele de donn´ees Kimball suit une approche ascendante pour entrepoˆt de donn´ees conception d’architecture dans laque- lle les magasins de donn´ees sont d’abord form´es en fonction des besoins de l’entreprise.

La m´ethodologie Kimball est assur´ee les phases suivantes:

* Collection des donn´ees.
* Int´egration des donn´ees.
* Analyse des donn´ees.
* Creation d’un rapport des donn´ees (reporting).

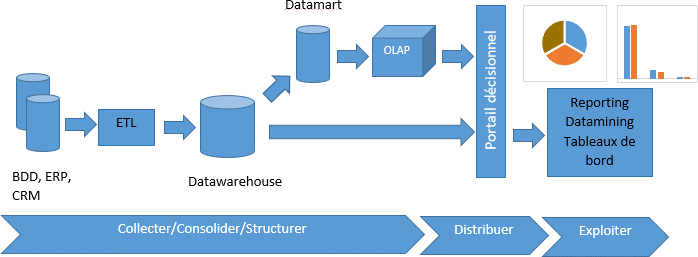


Figure 1: Cycle de vie d’un projet BI

# R´ealisation

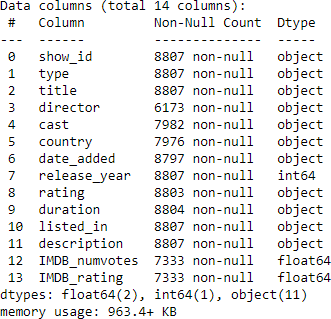
Comme nous l’avons dit plus tˆot, nous suivrons les ´etapes de la m´ethodologie kimball.

## Phase: Collection des donn´ees

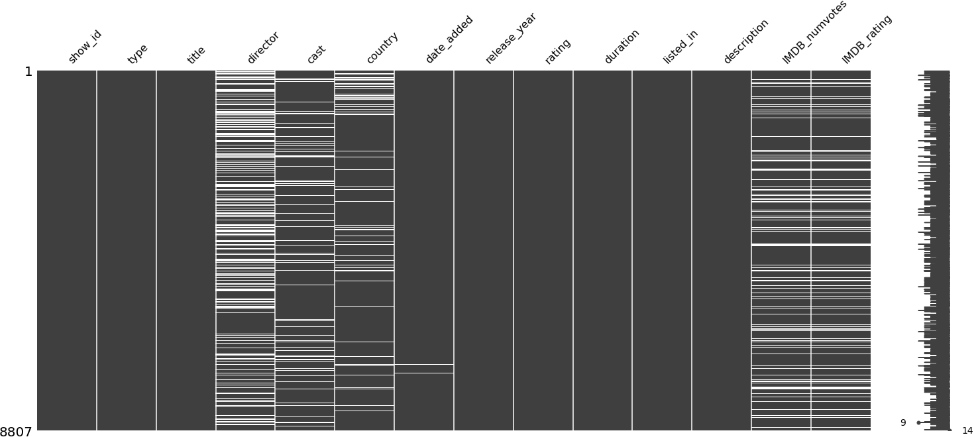
Nous utilisrons une base de donn´ees contient les titres (films et series) de entreprise multinationale am´ericaine **Netflix** et leurs notes sur IMDB.

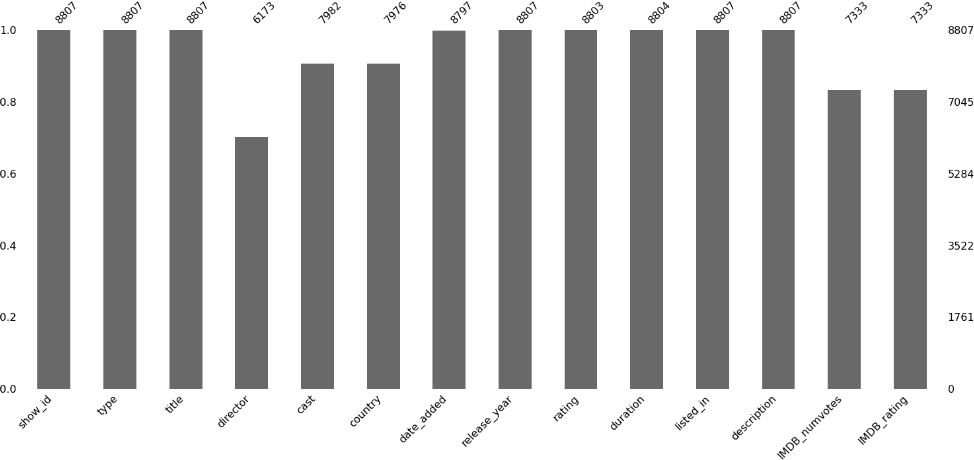


Notre base de donn´ees pr´esent se la forme d’une fictier text (txt) contient des informations concernant les titres.

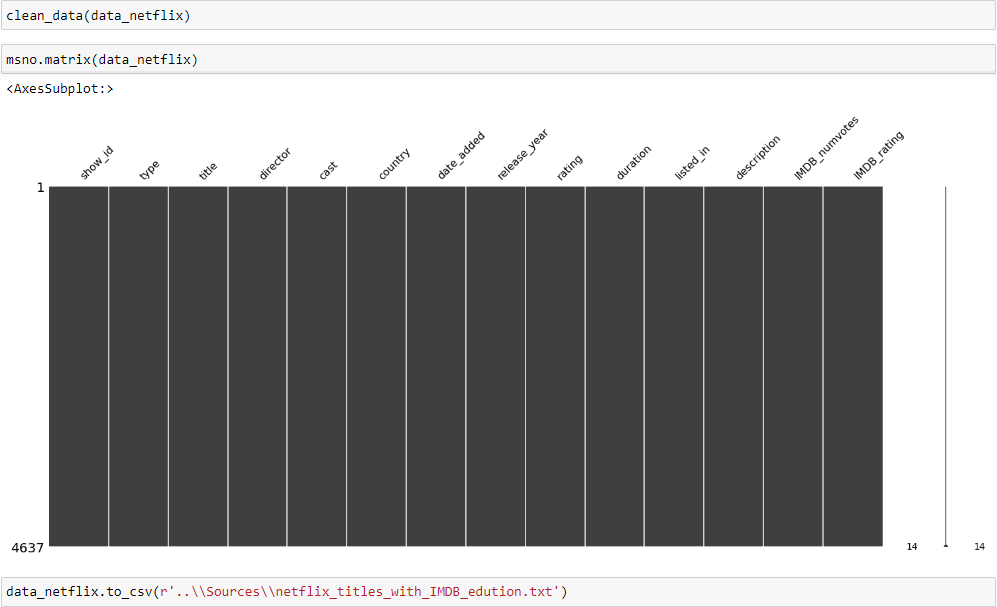


Lorsque nous voyons la figure pr´ec´edente nous observons que les nombres des cases non nulles dans les attributs ne sont pas ´egaux, c’est `a dire la pr´esence des valeurs manquantes.





La n´ecessit´e oblige a` nous d’´eliminer ces values manquantes pour ce l`a, nous allons utiliser un algorithme de nettoyage des donn´ees pour faire ¸ca.



## Int´egration des donn´ees

Cette ´etape consiste a` int´egrer et centraliser les donn´ees (datasets) un seul en- droit (Data warehouse) a` l’aide d’un service d’int´egration des donn´ees **SSIS** (**S**QL **S**erver **I**ntegration **S**ervices).

Pour faire c¸a, nous allons obliger a` faire la mod´elisation dimensionnelle ou` en va cr´eer le table des faits et les tables des dimensions. et de choisir un mod`ele (´etoile, flocon ou constellation) comme un mod`ele utilis´e pour lie les diff´erentes tables.

Dans cette partie, nous avons choisi d’utiliser le Mod`ele en ´etoile comme type de mod´elisation du Data Warehouse.

Notre mod`ele de conception compos´ee par 5 tables (4 tables de dimensions (”**Dim information**”, ”**Dim** **product**”, **”Dim** **date”** et ”**Dim** **rating**”) et un table des faits ”**Fact title**”).

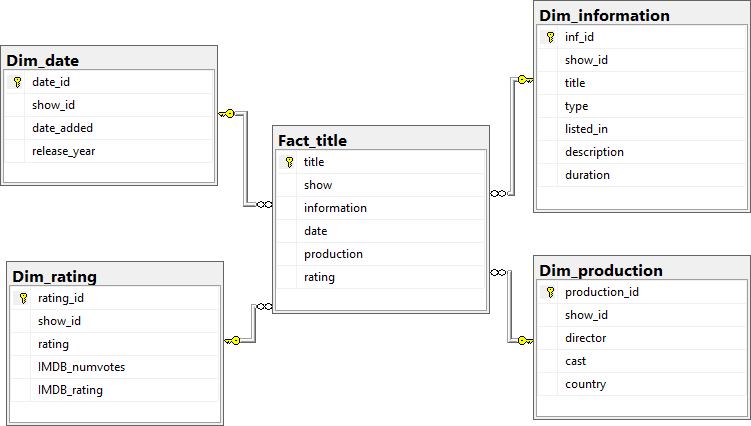
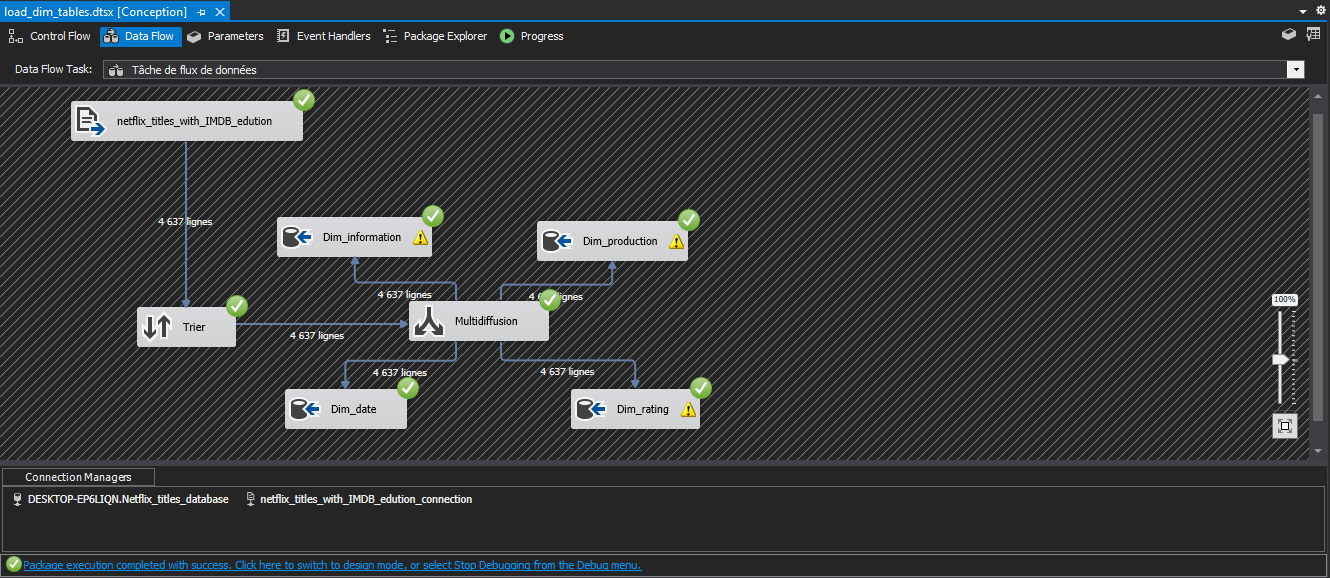
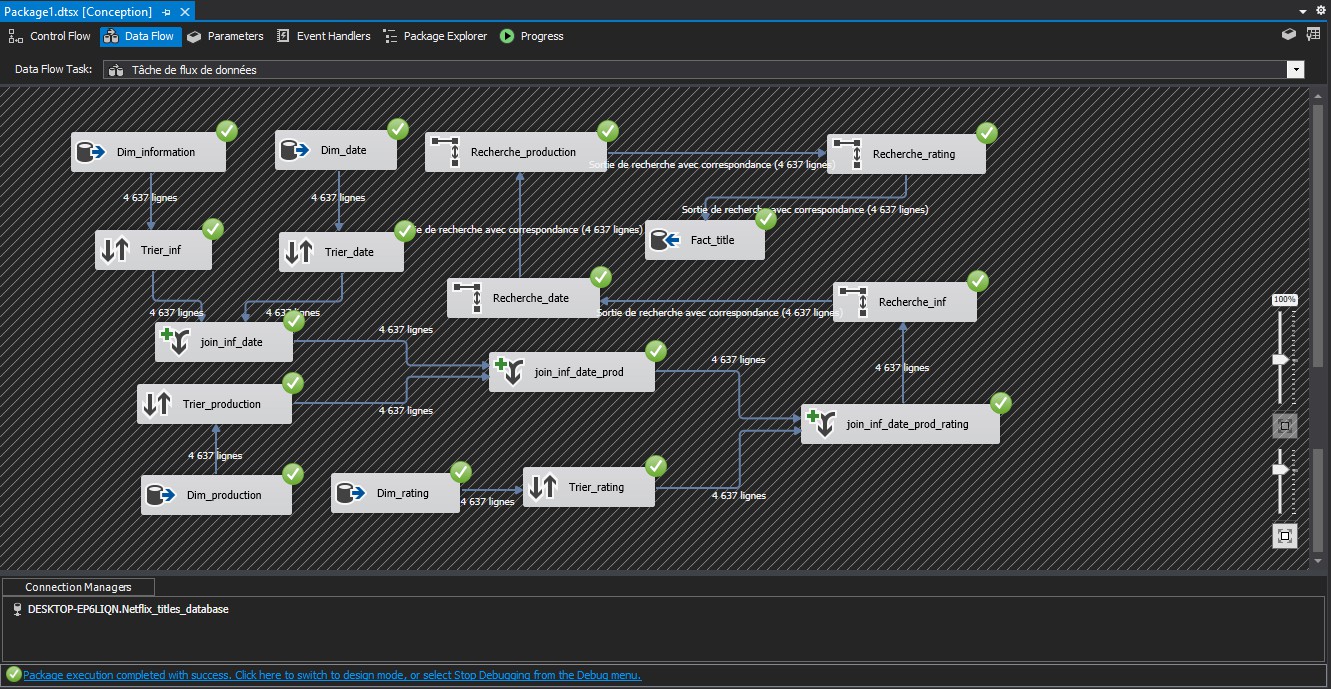


Figure 2: Conception du mod`ele physique du Data Warehouse

#### Chargement des tables des dimensions

****

#### Chargement de table des faits

****

## Analyse des donn´ees

Cette phase repr´esente la troisi`eme phase pour un projet ”BI” (phase de distribution des donn´ees), ou´ en va cr´eer notre cube OLAP a` fin d’analyser leurs donn´ees.

Cette processus se fait a` travers la **SSAS** (**S**QL **S**erver **A**nalysis **S**ervices).

#### Cr´eation d’un cube

Dans cette partie, nous allons d´etailler la cr´eation d’un cube a` partir d’un Data Warehouse.

La cr´eation d’un cube suit les etapes suivantes:

* + - * + **Data Source** (Source de donn´ees) : Cette ´etape consiste a` choisir la base de donn´ees concerne par le cube.
        + **Data View** (Vue de Source de donn´ees) : Cette ´etape consiste `a choisir a` partir de la base de donn´ees les tables concern´ees.
        + **Cube Structure** (structure de cube) : Cette ´etape consite `a montion- ner `a partir de quelle table nous ramenons les mesures bien ´evidemment notre cube.

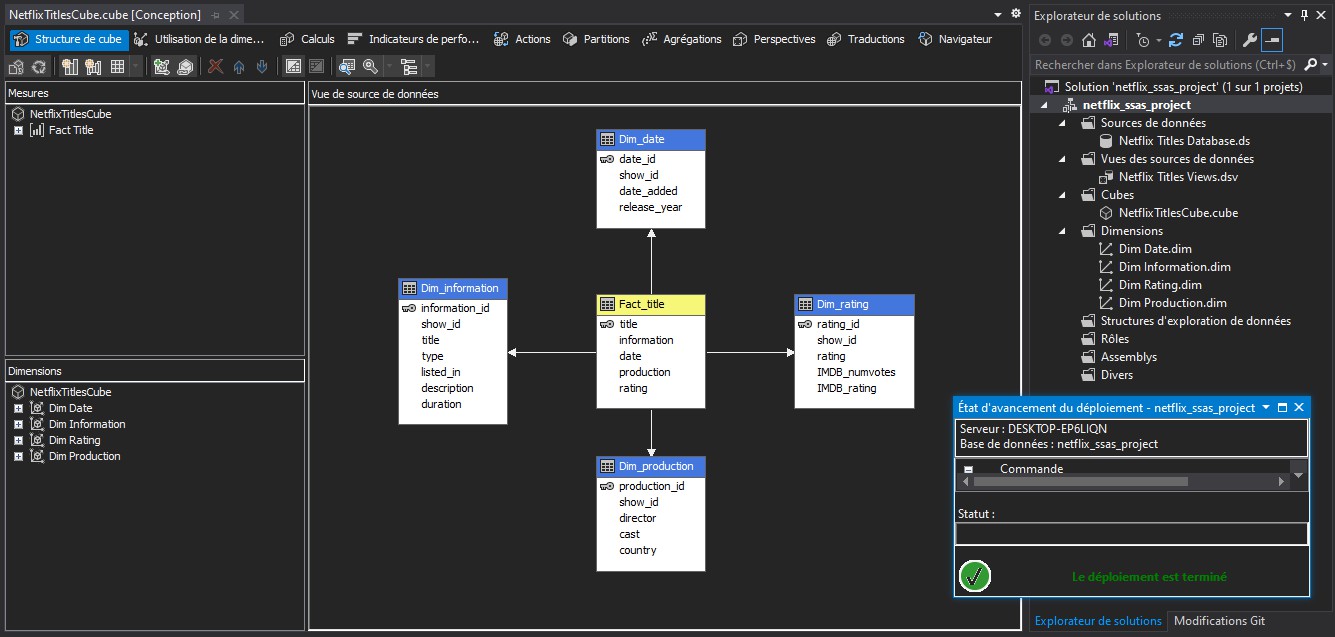


Figure 3: Cube OLAP: NetflixTitleCube

#### R´ecup´eration des donn´ee

La r´ecup´eration des donn´ees `a partir d’un cube se fait a` travers deux mani`eres soit par des requˆetes multidimensionnelles a` partir de l’utilisation **la langage ”MDX”**.

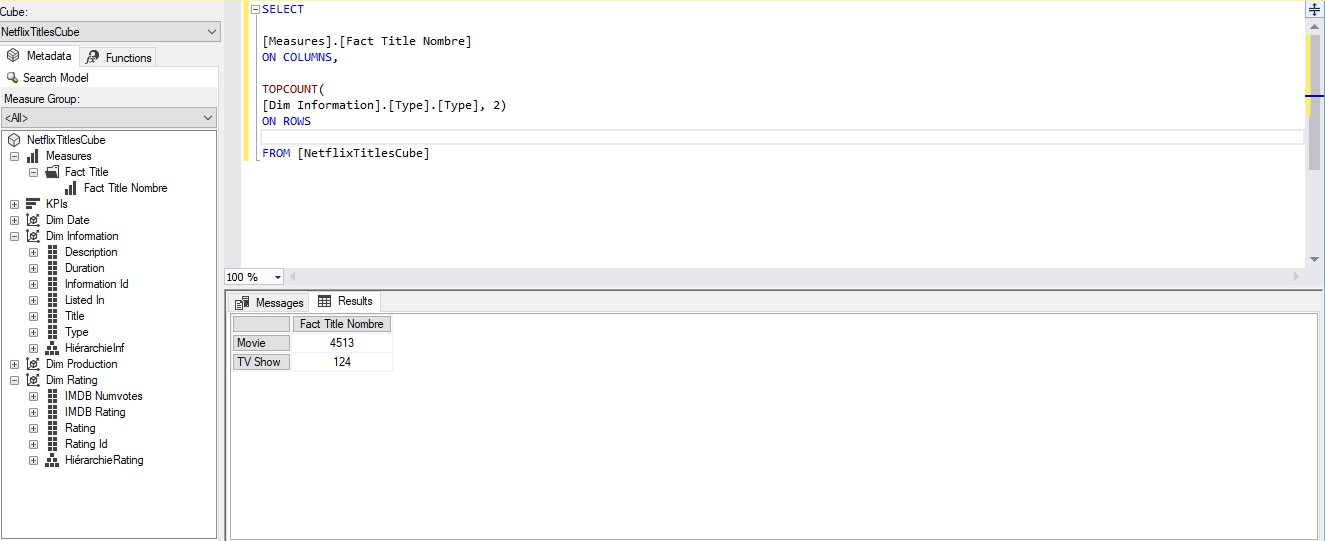


Figure 4: Requˆete MDX: Nombre des titre par type

Le figure ci-dessus montre qu’il y `a deux types des titre (les films et les serie TV) dans notre base de donn´ees. il dispose 4513 films et 124 series TV.

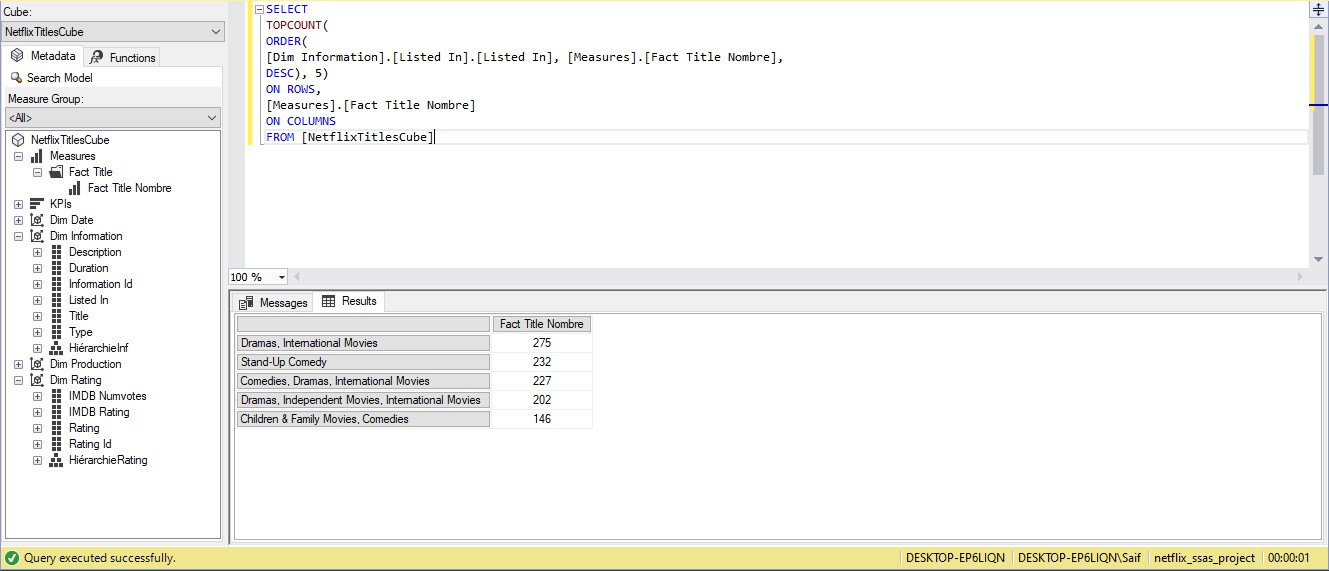


Figure 5: Requˆete MDX: Nombre des titre par genre

D’apr`es cette requˆete on observe que dans notre base de donn´ees les genres la plus nombreuse sont comme le r´esultat donne.

## Reporting

Cette phase consiste `a la cr´eation d’un rapport contient des informations

aident `a priser des d´ecisions `a travers les donn´ees r´ecup´er´ees a` cube.

partir du

Cette ´etape se fait a`

**S**ervices).

travers la service **SSRS** (**S**QL **S**erver **R**eporting

Notre Rapport **”NetflixTitleRatingReport”** est une repr´esentation graphique de nos donn´ees (se forme des graphes) faciles `a comprendre nous aider a` ex- tract des informations.

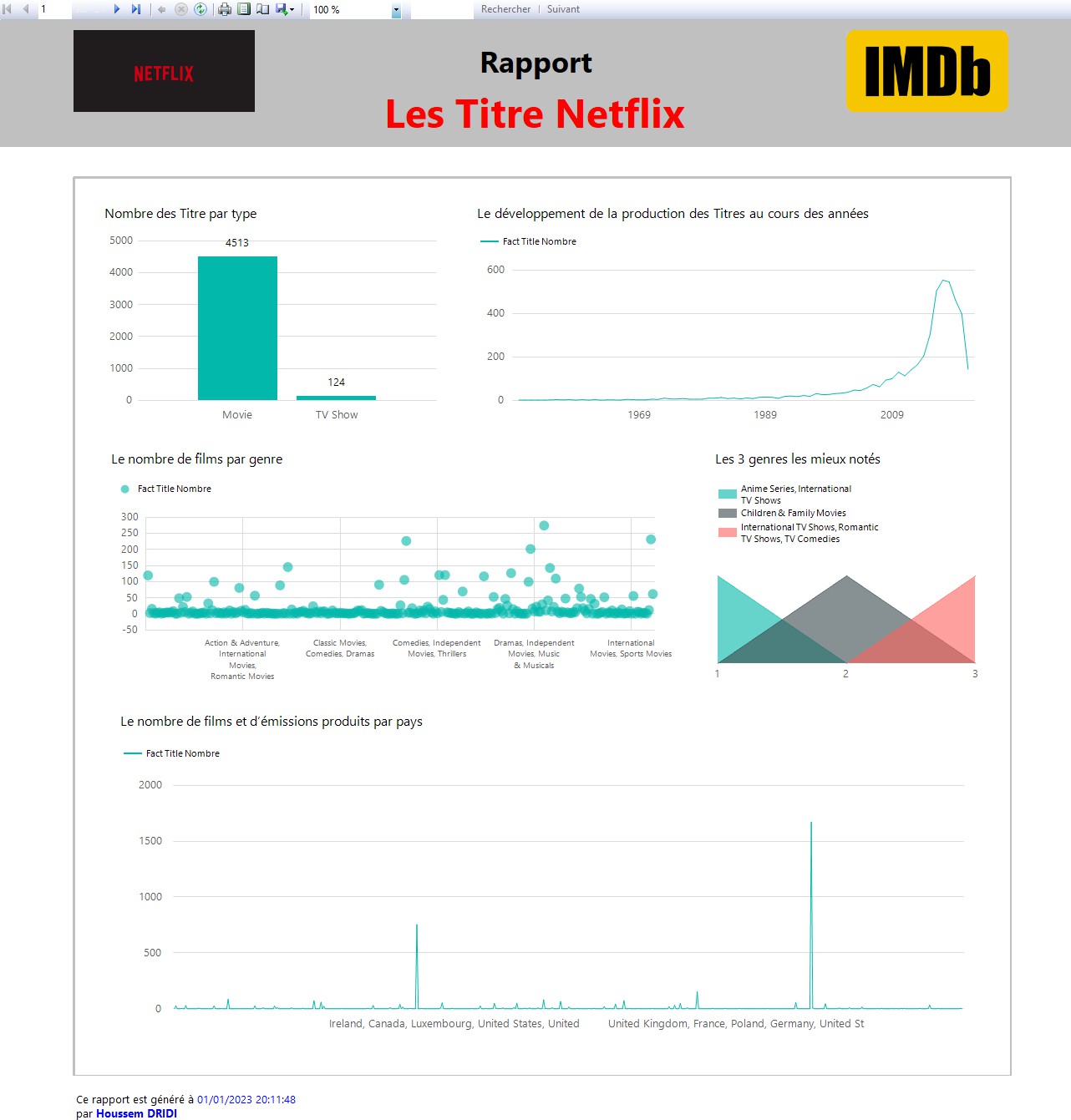


Figure 6: Le rapport: NetflixTitleRatingReport

# Conclusion

La data warehouse permet ainsi d’am´eliorer la prise de d´ecision et le d´eploiement de strat´egies plus efficaces. Il constitue un avantage concurrentiel important pour une entreprise.